



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Improving model prediction of soil N₂O emissions through Bayesian calibration

Citation for published version:

Myrgiotis, V, Williams, M, Topp, CFE & Rees, R 2017, 'Improving model prediction of soil N₂O emissions through Bayesian calibration', *Science of the Total Environment*.
<https://doi.org/10.1016/j.scitotenv.2017.12.202>

Digital Object Identifier (DOI):

[10.1016/j.scitotenv.2017.12.202](https://doi.org/10.1016/j.scitotenv.2017.12.202)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Science of the Total Environment

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Improving model prediction of soil N₂O emissions through Bayesian calibration

Vasileios Myrgiotis^{a,b}, Mathew Williams^b, Cairistiona F. E. Topp^a, Robert M. Rees^a

^a*SRUC, Edinburgh EH9 3JG, UK*

^b*School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JN, UK*

Abstract

The biogeochemical processes that lead to the production of N₂O in arable soils are controlled by temporally and spatially varying drivers. The need for prediction of soil N₂O emissions across scales means that agroecosystem biogeochemistry models are widely used to simulate N₂O emissions. Due to the parameter-dense nature of agroecosystem models their parameters have to be calibrated according to the soil and climatic conditions of the intended area of application. Bayesian calibration is considered one of the most advanced ways to complete this task. In this study, we calibrate nine parameters of the Landscape-DNDC process-based agroecosystem model, which are key to its N₂O prediction. The Metropolis-Hastings algorithm is used at four separate implementations in order to estimate parameter posterior distributions at four arable sites in the UK. The results of this process are visualised, summarised and assessed against measured N₂O data from ten independent arable sites. The study shows that, in many cases, soil N₂O emission peaks that were not predicted with the default model parameters were predicted

after calibration. Overall, the prediction of soil N₂O fluxes across all the sites that were considered was improved by 33% when using the calibrated parameters.

Keywords: N₂O, modelling, Landscape-DNDC, UK croplands, Bayesian calibration, Metropolis-Hastings

Highlights

- Nine key parameters of the Landscape-DNDC model were adjusted to UK conditions using a Bayesian algorithm and measured soil N₂O data
- The posterior parameter distributions led to a 56% improvement in the level-of-fit between measured and simulated soil N₂O data at the four calibration sites
- A 26% improvement in N₂O prediction was observed at 10 independent evaluation sites
- Calibration caused significant change in the suggested value for one parameter, a medium change for five parameters and a small change for the remaining three parameters

1. Introduction

The emission of nitrous oxide (N₂O) from agricultural soils is a particularly important side effect of modern agriculture because it is both a very powerful greenhouse gas (GHG) and a major contributor to ozone depletion. Nitrous oxide is produced in soils through the microbe-mediated processes of nitrification and denitrification. These processes are driven by the addition of nitrogenous fertiliser to the soil and controlled by micro-scale environmental and soil conditions (Butterbach-Bahl and Dannenmann, 2011). Consequently, field measurements of soil N₂O emissions include large uncertainties,

which are caused by the spatial and temporal variability of the fluxes (Rees et al., 2013; Cowan et al., 2014). Field measured N_2O data are used in various kinds of studies among which are those that focus on the development and evaluation of agroecosystem biogeochemistry (BGC) models.

Agroecosystem BGC models are scientific tools that mathematically represent the structure and underpinning biogeochemical processes in agroecosystems (Wainwright and Mulligan, 2005). They are used to simulate, among other things, N_2O emissions from soils at various spatial and temporal scales. Due to the high cost (time, labour, money) of field measurements of soil N_2O emissions and the large variability of emissions agroecosystem BGC models are widely used to quantify the N_2O footprint of agricultural ecosystems and to explore different methods to reduce it. The outputs of model simulations are affected by uncertainties around the model’s input data, its structure and its parameters (Wang and Chen, 2012). The impact of model input uncertainties on model outputs is often quantified in relative studies but the impact of model parametric uncertainty is less studied while the role of structural uncertainty is rarely examined (van Oijen et al., 2011; Rafique et al., 2015; Van Oijen et al., 2005; Lehuger et al., 2009; van Oijen et al., 2013).

Model parameters control various mathematically-represented processes and they can vary between locations (Li et al., 2015). This fact creates the need to adapt a model’s parameters to the edaphoclimatic conditions of its area of application. Through calibration the distributions of the parameters

of a model are updated based on information contained in measured data from locations which are representative of the area of interest (i.e. simulated area). In this way, the robustness of the model and the quality of its predictions is improved (Wainwright and Mulligan, 2005). Calibration is a process that includes iteration (i.e. repeated model evaluations) using parameter vectors (i.e. combinations of parameter values) that are sampled from the possible parameter space. Calibration algorithms are used to perform the necessary tasks by completing a set of steps that typically includes: (1) sampling for values from user-defined parameter-specific distributions; (2) executing the model code; (3) collecting the output of interest; (4) estimating the level-of-fit between simulated and measured data and (5) identifying the joint parameter distribution that produces an acceptable level-of-fit. Through this process, at each model iteration a new piece of information about the model (i.e. an estimate of model fit) is created.

It is the repetitive production of new knowledge on model behaviour that makes the Bayes theorem particularly useful in relation to model calibration. The Bayes theorem uses likelihood to relate any prior knowledge on an event to the posterior knowledge about it (Lehuger et al., 2009; van Oijen et al., 2013). In calibration terms, the Bayes formula is used to update the distribution of model parameters (posterior) by combining existing knowledge on the distribution of those parameters (prior) with a measure of how good the simulated and measured data fit with each other (likelihood) (Liepe et al., 2014; Lenormand et al., 2013; Gelman et al., 2014). This relationship can be

formally written as:

$$p(\theta|D) \propto l(\theta|D)p(\theta)$$

where θ is the vector of parameters, D the measured data and l the likelihood function. Various studies have used Bayesian methods with ecosystem BGC models and some of them have considered soil N_2O fluxes (Li et al., 2015; Rahn et al., 2012).

In this study we use a Bayesian calibration algorithm (i.e. Metropolis-Hastings) together with measured data for soil N_2O in order to update the distribution of selected parameters of the Landscape-DNDC model so that they better reflect UK edaphoclimatic conditions (Wang and Chen, 2012; Haas et al., 2012). Landscape-DNDC is a process-based ecosystem BGC model that uses inputs on soil, climate and management to predict flows of nutrient, energy and water in agricultural, forest and grassland ecosystems (Haas et al., 2012).

Two factors play an important role in model calibration especially with regards to soil N_2O emissions. Firstly, commonly used measures of goodness-of-fit (e.g. correlation coefficient, mean squared error) do not consider the role of uncertainty in measured data and time lags between measured and simulated data, which the nature of soil N_2O fluxes can amplify. In this context, Myrgeiotis et al. (2016) presented an algorithmic model evaluation method that considers the role of uncertainty in data and possible time lags. This

algorithm is used in study to quantify the goodness-of-fit between measured and simulated N_2O at each iteration. Secondly, process-based models, like Landscape-DNDC, depend on a large set of parameters. This fact can greatly increase the number of iterations that are needed to complete model calibration. For this reason, the use of sensitivity analysis to filter out the most important model parameters before proceeding to calibration is recommended. In the context of Landscape-DNDC, Myrgeiotis et al. (Submitted) have performed extensive analysis of the sensitivity of various model outputs (including N_2O) to the model's parameters and found that 9 parameters play key role to soil N_2O prediction.

In addition to how this study dealt with the aforementioned key issues it should be noted that our calibration is based on the separate execution of the Metropolis-Hastings algorithm at four different UK arable sites. This approach is used in order to ensure that no single site can affect the resulting parameter distribution disproportionately. Also, the climatic and soil-related variation that exists across the four sites is better integrated in the calibration process and their impacts are more strongly imprinted in the calibrated parameter distributions.

2. Materials and methods

The Metropolis-Hastings Bayesian calibration algorithm is used in this study to estimate the joint probability distribution for the most important soil biogeochemistry parameters of the Landscape-DNDC model that are

relevant to the prediction of soil N_2O fluxes. A sequence of processes was followed to estimate the joint probability distributions (Fig. 1). At a first stage, the prior values for 9 soil BGC parameters (key to N_2O prediction) are used to implement the algorithm separately at four UK arable sites (S1,S2,S3 and S4). Through the consideration of "evidence", which is represented by the degree of fit between measured and simulated data, the algorithm creates "informed" posterior distributions for the examined parameters. After the completion of this first stage the four sets of posterior distributions that were produced were merged into a single data set (i.e. parameters posterior). We then sampled for parameters from the posterior distributions and simulate soil N_2O fluxes at 10 independent UK arable sites (S5-S14). Based on the level-of-fit between measured and simulated N_2O we assess the success of the Bayesian calibration process in adapting the model's soil BGC parameters to UK edaphoclimatic conditions.

The tasks involved in the application of the Bayesian algorithm (i.e. creation of a Markov Chain, sampling etc) were performed using the Python PyMC library (Patil et al., 2010). As a preliminary step we completed testing evaluations of the calibration algorithm in order to define the number of iterations (i.e Markov Chain Monte Carlo steps) that are needed to achieve satisfactory chain convergence at each site. Based on these tests we chose to use 130000 iterations in total, of which 70000 were used for the burn-in phase (i.e. simulations intended to identify starting point for the Markov Chain). In order to evaluate the results of the Bayesian calibration, we run the model

500 times while sampling for parameters from the posterior distributions and we visualise the resulting simulated soil N₂O time series (mean and 95% confidence intervals) at 10 independent model evaluation arable sites (all located in the UK) as well as the four calibration sites.

2.1. Model performance metric

Model performance metrics such as the correlation coefficient (r), the mean squared error (MSE) and the root mean squared error (RMSE) are commonly used to describe the level-of-fit between measured and simulated values (i.e. summary statistics) during the calibration of ecosystem BGC models. However, the use of these statistical metrics with soil N₂O data can have unwanted side effects as a result of (1) time lags between measured and simulated data points and the (2) lack of consideration for the role of uncertainty in the measured data. In order to consider these aspects, we use the *accuracy* measure that is calculated by the model evaluation algorithm presented by Myrgeiotis et al. (2016), as the summary statistic. The accuracy measure quantifies the percentage of simulated data points that are within the corresponding measured ranges during a user-defined time period (e.g. 2 weeks, 1 year etc). At the same time the algorithm can consider how much a user-defined time lag can affect (i.e. increase) the estimated accuracy. In this study the impact of a 3-day time lag was considered by the model evaluation algorithm.

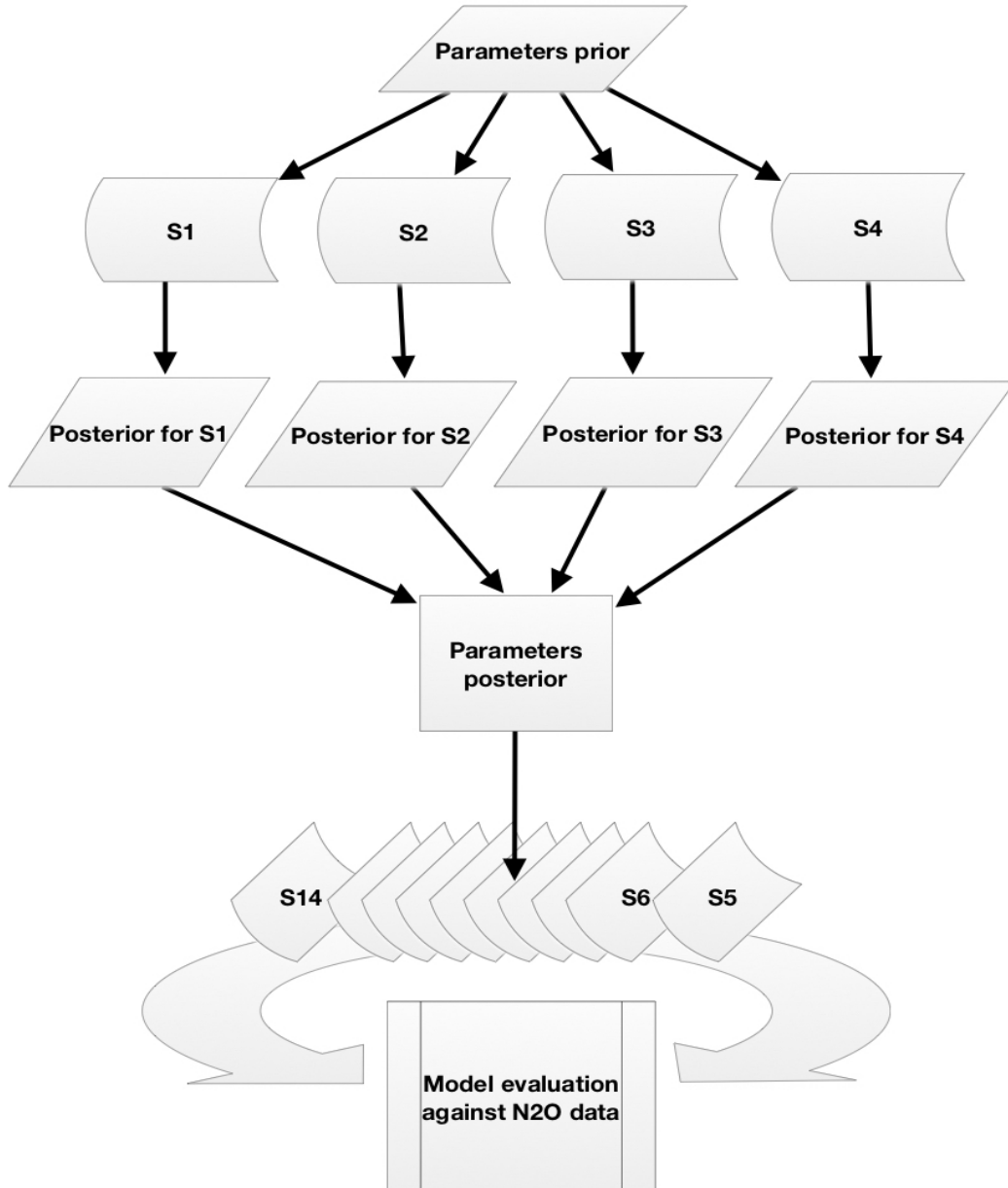


Figure 1: Graphic description of the calibration process. S1-S4 represent the four calibration sites and S5-S14 represent the 10 evaluation sites. Arrows represent flows of data/information. Leaning rectangles represent the distributions of the 9 parameters examined. The parameters posterior distributions that are derived from the calibration simulations for sites S1-S4 (calibration sites) are merged into a single "parameters posterior", which is then used to simulate sites S5-S14 (evaluation sites) and evaluate the model prediction of soil N₂O fluxes.

2.2. Convergence checking

The stabilisation of the distribution (i.e. convergence of the MCMC chain) for each model parameter examined was assessed visually (Sarrazin et al., 2016). This assessment was done by (1) extracting subsets of the estimated parameters at sets of 500 steps (i.e. model evaluations); (2) plotting their mean and median and (3) checking these plots for asymptotic behaviour (i.e. stability of mean and median value during a certain period).

2.3. Evaluation of posterior distributions

In order to assess the improvement in model soil N_2O prediction that is achieved when using the posterior distributions we ran the model at the four UK arable sites (i.e. calibration sites) and visualise the resulting N_2O time series. As a way to provide a quantitative evaluation of the fit between measured and simulated soil N_2O we also calculate the fraction of measured daily N_2O value ranges that overlap with the corresponding simulated daily N_2O ranges (defined by 95% confidence intervals) at each site. Because the posterior distributions are created based on the application of the Metropolis-Hastings algorithm at the four calibration sites it is important to examine whether the posterior distributions improve model performance at independent arable sites. For this reason, we also ran the model at 10 independent UK arable sites (i.e. evaluation sites) while sampling for parameter values from the posterior distribution. The model performance at these evaluation sites was assessed visually (i.e. plotting measured and simulated time series)

and quantitatively (i.e. using the aforementioned percentage-of-overlap approach).

2.4. Landscape-DNDC

Landscape-DNDC is a process-based ecosystem BGC model that can simulate nutrient, energy and water flows in arable, grassland and forest ecosystems (Haas et al., 2012). It belongs to a group of models that are based on the DeNitrification-DeComposition (DNDC) model (Gillespy et al., 2014). The model has a modular structure with five different modules simulating certain parts of the ecosystem: (1) management, (2) plant growth, (3) water cycling, (4) soil microclimate and (5) soil biogeochemistry. Nitrous oxide is modelled as being the result of the processes of nitrification and denitrification using the hole-in-the-pipe approach (Davidson et al., 2000). The activation of each process (nitrification/denitrification) is controlled by (1) the oxygen content (i.e. anaerobic conditions), (2) the soil's pH and temperature and (3) the soil's N substrate. The soil BGC module of the model depends on 123 parameters; excluding parameters whose value is set to 0 by default. The sensitivity of simulated N_2O to these parameters has been quantified by (Myrriotis et al., Submitted). For this study, we used 9 parameters that were found to be the most important for the prediction of N_2O as well as of the precursing NO and NO_2^- (Table 1). The lower and upper boundary for each parameter are predefined by the model's developers and cannot be exceeded while their distribution is uniform.

2.5. Datasets

Field-scale data on model inputs (i.e. climate, soil texture, pH, bulk density, C and clay content) and soil N₂O fluxes (kgNha⁻¹) are used in order to run and calibrate the model (from Sylvester-Bradley et al. 2015 and the Agricultural and Environmental Data Archive). Separate data sets were used in model calibration (Table 2) and model evaluation (Table 3). These ancillary soil data come from analyses of soil samples that were collected at Gleadthorpe (Nottinghamshire), Edinburgh (Scotland), Boxworth (Cambridgeshire) and Terrington (Norfolk) while the corresponding climate data come from on-field weather stations. Nitrous oxide samples were collected using static flux chambers that were regularly positioned in the measurement area (Hensen et al., 2013; Saggar et al., 2010).

Table 1: Landscape-DNDC parameters that were selected for calibration

Description	Name	Lower Boundary	Upper Boundary
Microbial death rate	AMAXX	0.9817	1.6362
Microbial denitrifier fraction	DENIFRAC	0.5250	0.8750
Fraction of decomposed carbon that goes to the dissolved organic carbon pool	EFFAC	0.5250	0.8750
Reduction constant for N ₂ O diffusion	D N ₂ O	0.0465	0.0775
Reduction constant for NO diffusion	D NO	0.0547	0.0912
Microbial efficiency for NO ₂ denitrification	EFF NO ₂	0.3210	0.5350
Reaction rate for nitrification	KNIT	0.7500	1.2500
Microbial growth rate	MUEMAX	3.6547	6.0913
Microbial growth rate for denitrification on NO ₂	MUE NO ₂	0.5025	0.8375

Table 2: Information on sites used in model calibration

Crop type	Code	Reference	Location	Simulated years	Soil texture (g/cm3)	BD (%)	Clay content	pH (%)	Soil C (%)	Maximum WFPS (%)	Minimum WFPS (mm)	Total precipitation (kgN/ha)	Fertiliser
Spring oilseed rape	S1	Sylvester-Bradley et al. (2015)	Edinburgh	2007-2010	SNCL	1.27	36	6.2	5	49	11	807	96
Winter wheat	S2	Sylvester-Bradley et al. (2015)	Boxworth	2007-2010	CLLO	1.43	34	7.6	1.97	92	25	502	200
Winter wheat	S3	Smith et al. (2012)	Boxworth	2002-2005	CLAY	1.2	50	8.2	2.4	56	34	543	160
Winter wheat	S4	Thorman et al. (2017)	Terrington	2001-2004	SLCL	1.38	32	8.1	1.67	65	34	759	220

Notes: Year of experiment is the last simulated year; Total precipitation refers to year of experiment only; SALO: Sandy Loam; SNCL: Sandy Clay Loam; CLLO: Clay Loam

Table 3: Information on sites used in model evaluation

Crop type	Code	Reference	Location	Simulated years	Soil texture (g/cm3)	BD (%)	Clay content	pH (%)	Soil C (%)	Maximum WFPS (%)	Minimum WFPS (mm)	Total precipitation (kgN/ha)	Fertiliser
Winter wheat	S5	Sylvester-Bradley et al. (2015)	Terrington	2008-2011	SALO	1.35	11	8.27	1.84	52	16	472	180
Winter wheat	S6	Sylvester-Bradley et al. (2015)	Boxworth	2008-2011	CLLO	1.43	32	7.5	1.5	85	21	378	200
Winter wheat	S7	Sylvester-Bradley et al. (2015)	Boxworth	2009-2012	CLAY	1.08	49	8.2	2.7	66	32	730	200
Winter barley	S8	Sylvester-Bradley et al. (2015)	Gleadthorpe	2008-2011	SALO	1.33	9.7	6.3	1.5	49	10	379	160
Spring barley	S9	Sylvester-Bradley et al. (2015)	Gleadthorpe	2009-2012	SALO	1.33	11	6.0	1.6	56	18	888	140
Winter barley	S10	Sylvester-Bradley et al. (2015)	Edinburgh	2007-2010	SNCL	1.23	36	6.2	4	49	11	807	140
Spring barley	S11	Sylvester-Bradley et al. (2015)	Edinburgh	2008-2011	CLLO	1.22	34	6.7	4.9	51	33	1312	144
Winter oilseed rape	S12	Sylvester-Bradley et al. (2015)	Edinburgh	2008-2011	CLLO	1.26	34	6.7	4.9	62	33	1312	160
Winter wheat	S13	Sylvester-Bradley et al. (2015)	Edinburgh	2007-2010	SNCL	1.26	36	6.2	4	55	29	807	148
Winter wheat	S14	Sylvester-Bradley et al. (2015)	Edinburgh	2008-2011	CLLO	1.26	34	6.7	4.9	58	31	1312	180

Notes: Year of experiment is the last simulated year; Total precipitation refers to year of experiment only; SALO: Sandy Loam; SNCL: Sandy Clay Loam; CLLO: Clay Loam

3. Results

3.1. Estimated posterior distribution

The visual assessment of the convergence of the MCMC chain showed that, in most cases, the distribution stabilised after 30000 to 40000 post burn-in steps (i.e. model runs). However, the distribution for the EFFAC, MUEMAX and AMAXX parameters, at the S1 site, stabilised after 50000 simulations. By merging the results of the Bayesian calibration for each of the four sites we produced Figure 2, which presents the posterior distribution for the 9 parameters. Table 4 summarises the results of the calibration and presents the suggested, lower and upper value (i.e range) of the prior distribution next to the respective values from the posterior distribution. For the posterior distribution the suggested value is either the mode or median of the distribution, depending to the shape of the respective posterior distribution, while for the prior it is the default value. For the posterior distribution the presented range is 95% confidence interval (CI) of the data while for the prior distribution they are the default min/max limits as set by the model's developers (i.e. uniform distributions).

The improvement in the parameter ranges for the 9 examined model parameters appears to be rather small (Table 4). This is explained by the fact that the posterior parameter ranges presented are in essence the 95% CIs of mainly skewed distributions (positively and negatively). Because of that the CI is not adequate to describe the knowledge that was gained by the Bayesian calibration processes. In this respect, it is the shape of the posterior

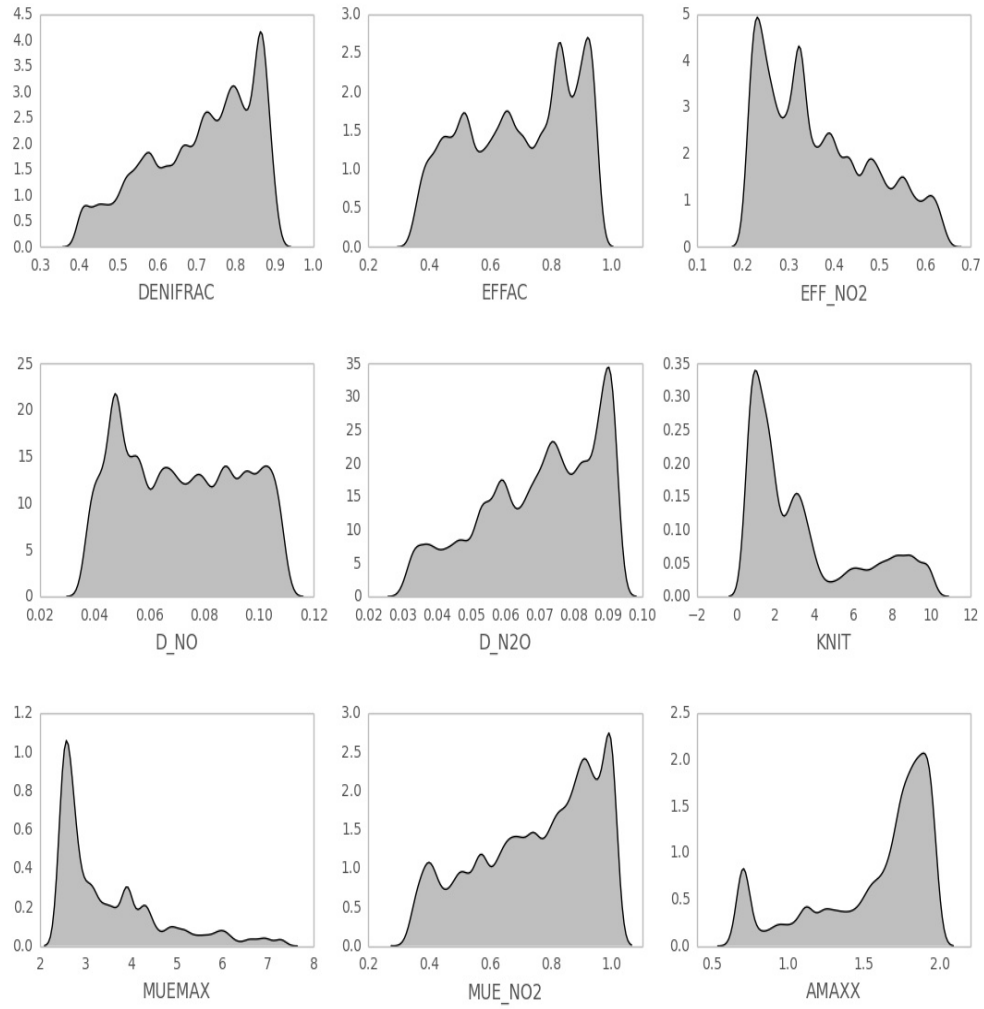


Figure 2: Kernel density estimations for the posterior distributions of the nine model parameters

distribution (Figure 2) for each parameter that can be more informative. In order to make this more clear we added the prior and posterior suggested values in Table 4 along with a reference to the central tendency measure (i.e. mean, median, mode) that appears as best suited to describe the shape of the posterior distribution. We also estimated the level of change in the suggested value for each parameter pre and post calibration. The level of change is expressed as a percentage and shows how much did calibration alter the suggested value relative to its default value. It should be noted that for most of the parameters the posterior distribution appears to be abruptly cut at the edge (left or right) of the range of possible values. This happens because the parameters of Lanscape-DNDC are allowed to have values that are within certain limits. In this context, the results presented in Figure 2 suggest that the predefined parameter limits might be too narrow for some of the parameters examined (i.e. KNIT, MUEMAX, AMAXX).

The mean level of change (%) that was achieved for all nine parameters was 29%. Calibration has led to a small level of change ($< 10\%$) of the suggested value for three parameters (i.e. DENIFRAC, EFFAC, D NO). A medium level of change (between 10% and 40%) was achieved for the majority of parameters (i.e. AMAXX, EFF NO2, D N2O, MUEMAX, MUE NO2) while one parameter was strongly (changed by 125%) affected by the calibration (i.e. KNIT). KNIT is a key parameter that controls the reaction rate for nitrification and has a strong impact not only on the simulated soil N₂O fluxes but also on the production and emission of N₂ and NO. The

reduction constant for N₂O diffusion (D N₂O) is another gas diffusion-related parameter which affects how much N₂O is diffused through the soil. The remaining four parameters for which a medium level of change was estimated (i.e. AMAXX, EFF NO₂, MUEMAX, MUE NO₂) are parameters related to the dynamics of the soil’s microbial population.

Table 4: Summary of prior and posterior parameter values (suggested values and respective ranges) for the examined model parameters

Parameter	Prior		Posterior		Level of change (%)
	range	suggested	range	suggested central tendency	
AMAXX	0.6000 - 1.9600	1.3090	0.7100 - 1.9500	1.7200	median 31
DENIFRAC	0.4000 - 0.9000	0.7000	0.4600 - 0.8800	0.7300	mode 4
EFFAC	0.3500 - 0.9500	0.7000	0.4000 - 0.9300	0.7100	mode 1
EFF NO ₂	0.2140 - 0.6420	0.4280	0.2200 - 0.5900	0.3400	mode -21
D NO	0.0365 - 0.1095	0.0730	0.0460 - 0.1050	0.0709	median/mean -3
D N ₂ O	0.0310 - 0.0930	0.0620	0.0380 - 0.0910	0.0730	mode 18
KNIT	0.5000 - 10.000	1.0000	0.6000 - 9.0000	2.2500	mode 125
MUEMAX	2.4360 - 7.0390	4.8730	2.4700 - 6.0000	3.0700	median -37
MUE NO ₂	0.3350 - 1.0050	0.6700	0.3950 - 1.0000	0.7900	median 18

The posterior parameter distributions presented in Figure 2 are based on four sets of data (one for each calibration site) that were merged into a single dataset. However, we did not observe any stratification in the data (Fig 2) because the posterior distributions for the 9 parameters were very similar across the four sites and, therefore, the results of no single site dominate the final posterior distributions. Also, the amount of samples used in the calibration process combined with the small number of parameters examined meant that the chances of correlation between sampled parameters was low; this was confirmed by our analysis on the correlation between parameters.

3.2. Evaluation of the posterior distribution

The Bayesian calibration process improved the fit between measured and simulated data at the four calibration sites. In order to evaluate the posterior distributions we ran the model at a set of independent sites while using 500 random parameter vectors drawn from the posterior distributions. This number of simulations/samples was enough to achieve a stabilised distribution for the simulated soil N₂O. Figure 3 presents the simulated (with default parameters) and measured soil N₂O fluxes at a subset of four sites (out of 10 in total) that were used in model evaluation. Figure 4 presents the measured and simulated soil N₂O (including 95% CIs) at the same four sites (i.e. S5, S6, S8, S13) when the model was run using parameters sampled from the posterior distributions. The plots of simulated and measured soil N₂O under default and calibrated parameters for all 14 sites (4 calibration and 10 evaluation sites) that were used in this study are presented in the appendix.

A visual assessment of the results of the simulations (Figs 3 and 4) before and after calibration shows that the improvement in the level-of-fit between measured and simulated N₂O has been significant. This improvement in fit can be seen for all 14 sites used in this study. Interestingly, N₂O emission peaks that did not exist in the results of the model when using default parameters appear under the calibrated model simulations. However, there are sites for which the calibrated model, just like the default model, was unable to capture peaks in soil N₂O emissions (e.g. S9, S14). This might suggest that the processes, which caused the emission peaks at those sites, are not

well represented in the model. It should be added at this point that the patterns of N_2O emissions at some evaluation sites (e.g. S10, S11, S13) are difficult to explain, and simulate, due to large delays between fertiliser application (i.e. happened between julian day 70 and 150 in our sites) and emission peaks.

In order to add a quantitative aspect to the assessment of the model's predictive accuracy for the simulations with the calibrated model we quantified the percentage of measured data points (including their standard error) that lie within the 95% CIs of the corresponding simulated data (i.e. Observed-vs-Simulated fit). Our results (Table 5) show that, on average across all 14 sites, 70% of measured data lied within the respective simulated CIs. For the majority of the simulated sites, the use of calibrated parameters led to noticeable improvements in the prediction of patterns and magnitudes in measured soil N_2O data (Figs 4 and 3). Overall, model calibration led to a 33% improvement in the level-of-fit between measured and simulated soil N_2O data across all sites with the improvement being 56% at the four calibration sites (S1-S4) and 26% at the 10 evaluation sites (S5-S14).

The 95% CI represents the limits within which the simulated N_2O lies (with a 95% possibility) when considering the uncertainty around the "true" value of 9 key model parameters. In order to quantify the level of uncertainty over predicted N_2O , as it is imprinted in the daily 95% CIs, we estimated the relative margin of error (for each day and site). The mean relative margin of error across all 14 sites considered in this study was 125%. Put in other

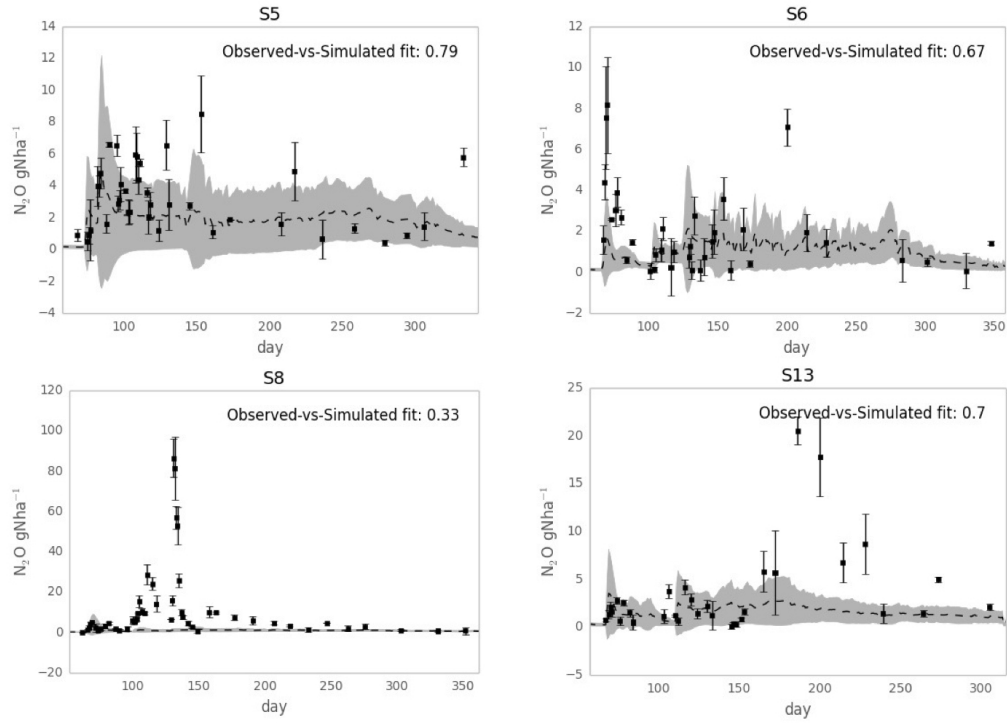


Figure 3: Plots of simulated (---) and measured (■) soil N_2O based on simulations with the default model parameters at four evaluation sites (S5, S6, S8, S13). The black dashed line (---) represents the mean simulated N_2O , the grey bandwidths represent the 95% CI, the square points (■) represent the mean measured N_2O and the error bars (in black) represent the corresponding standard error of the N_2O measurements.

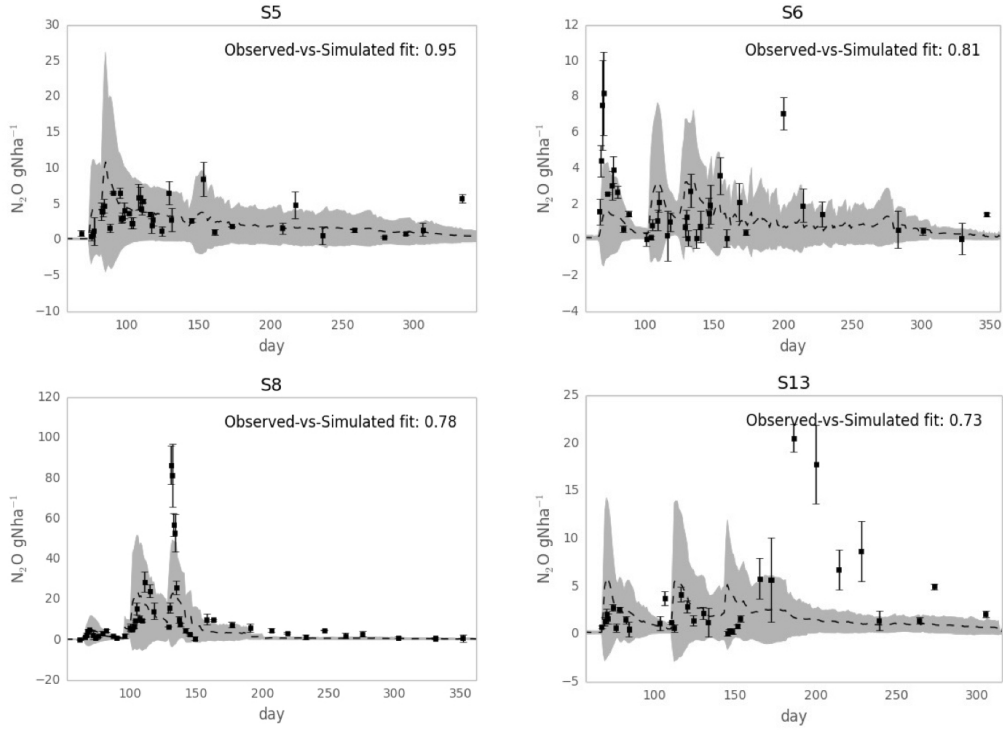


Figure 4: Plots of simulated (—) and measured (■) soil N_2O based on simulations with the posterior model parameters at four evaluation sites (S5, S6, S8, S13). The black dashed line (—) represents the mean simulated N_2O , the grey bandwidths represent the 95% CI, the square points (■) represent the mean measured N_2O and the error bars (in black) represent the corresponding standard error of the N_2O measurements.

Table 5: Level of fit between observed and simulated soil N₂O data.

Site code	Obs-vs-Sim fit	
	default	calibrated
1	0.60	0.64
2	0.61	0.76
3	0.38	0.74
4	0.16	0.59
5	0.79	0.95
6	0.67	0.81
7	0.30	0.63
8	0.33	0.78
9	0.62	0.62
10	0.55	0.65
11	0.60	0.64
12	0.66	0.76
13	0.70	0.73
14	0.62	0.76
Mean	0.54	0.72

words, the 95% CI for a daily simulated N₂O has a radius that is, on average, equal to 125% of the simulated mean daily N₂O.

4. Discussion

This study showed that the Bayesian approach to model calibration (through the Metropolis-Hastings algorithm) can be applied in order to calibrate the soil BGC module of a process-based agroecosystem model. We focused on a key agricultural greenhouse gas (i.e. N₂O) for which model-based prediction remains a rather difficult task due to measurement variability and knowledge gaps on the underlying processes.

The gross nitrification rate parameter in Landscape-DNDC (i.e. KNIT) was the parameter that was changed the most due to the calibration pro-

cess. Nitrification rate in soils can be measured but it is well known that various soil variables affect this rate (e.g. temperature, moisture, pH) and KNIT can therefore vary considerably between agricultural soils (Stange and Neue, 2009). Even though the simulated nitrification is adapted according to soil conditions it is proven by our study that this parameter needs special attention by users of Landscape-DNDC since it can greatly alter the way the model predicts patterns and magnitudes in soil N_2O emissions. In addition to that, among the parameters for which the estimated level of change (due to calibration) was between 10 and 30% (medium level of change) there were parameters that directly control the soil’s microbial population, which is known to also affect the soil’s nitrification rate (Stange and Neue, 2009).

The implementation of the Bayesian-based methodology that we decided to follow was straightforward and some of the problems that are common to Bayesian calibration have not affected this study. Such problems include (1) high correlations between sampled parameters, which is usually a result of including a large number of model parameters in the calibration process; (2) considerable number of simulations/steps needed to achieve stabilisation of the shape of the posterior distribution and (3) failure of the posterior distribution to provide parameter vectors that lead to good model performance at sites other than the calibration sites.

One important reason for having avoided these problems was the inclusion of a small set of parameters to the calibration process (i.e. 9 parameters). We based our decision on which parameters to select on a detailed global

parametric sensitivity analysis of the model’s most important outputs. This fact together with the improvement in the model’s accuracy at all the evaluation sites is a confirmation that we have included parameters key for N_2O prediction to the calibration. This does not mean that we have not excluded important parameters from the calibration but the high estimated mean relative margin or error is a proof of the importance of the parameters that were examined. Interestingly, only three of the soil BGC parameters used in this study were included in the only study (to our knowledge) that calibrated Landscape-DNDC against N_2O ; even though this was a forest-focused study (Rahn et al., 2012).

Another important aspect of this study is the implementation of the calibration algorithm, and all the associated tasks, separately at four UK arable sites (i.e. calibration sites), which is not a common practice. This setup assured that the posterior distribution is not influenced by the conditions (i.e. soil, climate) of a single site and is more representative of UK edaphoclimatic conditions as they are reflected in the ancillary data from the four calibration sites. In this respect, the larger the number of sites used in the calibration and evaluation process the better for the overall quality of the study.

In addition to the individual application of the calibration algorithm at the four calibration sites this study has used a novel model performance metric. We believe that the accuracy metric, provides a robust assessment of how well a model fits the measured data. The present study shows that the inclusion of the accuracy metric to a Bayesian calibration framework does

not affect the process negatively and we can reasonably assume that it can be used in the place of commonly used metrics like RMSE and r . This is particularly so in the case of calibration studies that use measured data with the characteristics of those on soil N_2O fluxes (i.e. sporadic measurements, high measurement uncertainty) for which the caveats of the use of RMSE and/or r are important (Myrgeiotis et al., 2016).

5. Conclusions

Nitrous oxide is produced in soils through a complex sequence of biogeochemical processes and this leads to uncertainty in measurements and complications in the modelling of N_2O emissions. Agroecosystem BGC models are diagnostic tools that allow us to predict emissions under various edapho-climatic and management conditions. Deliberate model application requires that their parameters are adjusted to the prevailing conditions at the area of application. In this study, we showed that this objective can be achieved through Bayesian calibration. We believe that the success of Bayesian calibration against soil N_2O depends strongly on how well the goodness-of-fit between measured and simulated data is incorporated in the calibration process. Also, the use of sensitivity analysis to identify which parameters are most important for the prediction of soil N_2O helps to reduce the number of parameters considerably and improves the robustness of the calibration process. Our calibration was successful in visibly improving predictions of N_2O at 14 UK sites. However, the mean relative margin or error around

the daily simulated N_2O was rather high. This fact suggests that even with updated parameter distributions the parametric sensitivity of the predicted N_2O is high. Also, the calibration process appeared to be limited by the fact that parameter values in Landscape-DNDC cannot exceed certain boundaries. Our results form an argument for the readjustment of parameter limits in Landscape-DNDC. It is important to note that the quality of calibration is strongly dependent on the size and quality of the field measured data use. We suggest that studies on model calibration should always try to use data from as many sites (within an edaphoclimatic area) as possible.

6. Acknowledgements

The authors are grateful to the Scottish Government for supporting this work and to the NERC Greenhouse Gas Emissions & Feedbacks program for its contribution. The data collection for the MIN-NO project was carried out by Rachel Thorman, Kate Smith, Roger Sylvester Bradley (ADAS) and Valentini Pappa (SRUC), supported by technical staff. The collection of data was sponsored by DEFRA and the Scottish Government through Sustainable Arable LINK Project LK09128, and we acknowledge the contributions of ADAS, Agricultural Industries Confederation, Bayer CropScience, British Sugar, Country Land and Business Association, The Co-operative Group, Frontier Agriculture, CF Fertilisers, AHDB cereals and oilseeds, Hill Court Farm Research, NFU, North Energy Associates, PGRO, Renewable Energy Association, Rothamsted Research (North Wyke), Scotch Whisky Research

Institute, SoilEssentials, SRUC, Vivergo fuels, Warburtons and Yara UK.

7. Appendix

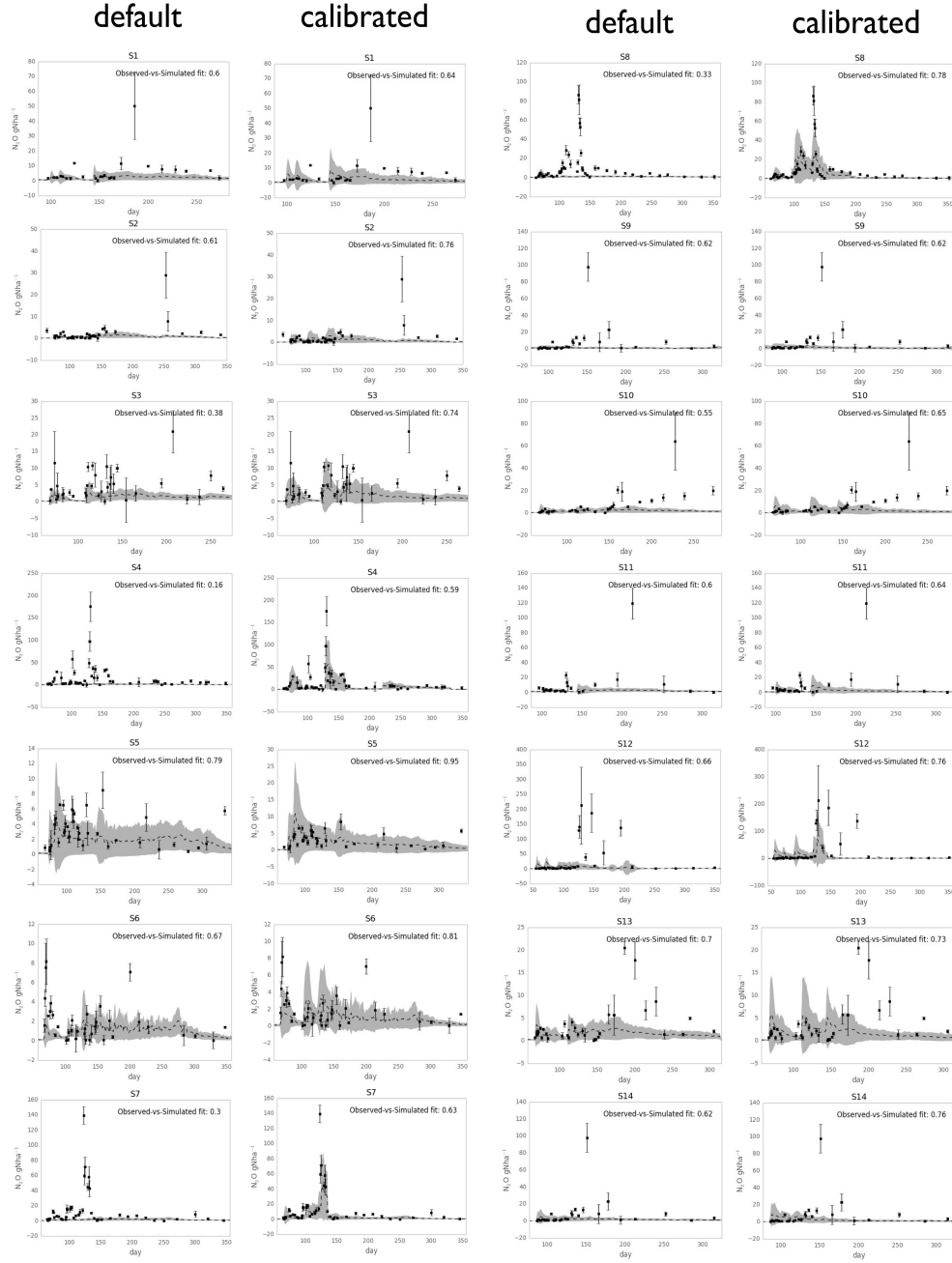


Figure 5: Plots of simulated (—) and measured (■) soil N_2O based on simulations with the default and calibrated model parameters. The black dashed line (—) represents the mean simulated N_2O , the grey bandwidths represent the 95% CI, the square points (■) represent the mean measured N_2O and the error bars (in black) represent the corresponding standard error of the N_2O measurements.

8. References

- Butterbach-Bahl, K. and Dannenmann, M. (2011), ‘Denitrification and associated soil N₂O emissions due to agricultural activities in a changing climate’, *Current Opinion in Environmental Sustainability* **3**(5), 389–395.
- Cowan, N. J., Famulari, D., Levy, P. E., Anderson, M., Bell, M. J., Rees, R. M., Reay, D. S. and Skiba, U. M. (2014), ‘An improved method for measuring soil N₂O fluxes using a quantum cascade laser with a dynamic chamber’, *European Journal of Soil Science* **65**(5), 643–652.
- Davidson, E. A., Keller, M., Erickson, H. E., Verchot, L. V. and Veldkamp, E. (2000), ‘Testing a Conceptual Model of Soil Emissions of Nitrous and Nitric Oxides: Using two functions based on soil nitrogen availability and soil water content, the hole-in-the-pipe model characterizes a large fraction of the observed variation of nitric oxide and nitrous oxide emissions from soils’, *Bioscience* **50**(8), 667–680.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2014), ‘Bayesian data analysis’.
- Gilhespy, S. L., Anthony, S., Cardenas, L., Chadwick, D., del Prado, A., Li, C., Misselbrook, T., Rees, R. M., Salas, W., Sanz-Cobena, A., Smith, P., Tilston, E. L., Topp, C. F. E., Vetter, S. and Yeluripati, J. B. (2014), ‘First 20 years of DNDC (DeNitrification DeComposition): Model evolution’, *Ecological Modelling* **292**.

- Haas, E., Klatt, S., Fröhlich, A., Kraft, P., Werner, C., Kiese, R., Grote, R., Breuer, L. and Butterbach-Bahl, K. (2012), ‘LandscapeDNDC: a process model for simulation of biosphere–atmosphere–hydrosphere exchange processes at site and regional scale’, **28**(4), 615–636.
- Hensen, A., Skiba, U. and Famulari, D. (2013), ‘Low cost and state of the art methods to measure nitrous oxide emissions’, *Environmental Research Letters* **8**(2), 025022.
- Lehuger, S., Gabrielle, B., Oijen, M. v., Makowski, D., Germon, J. C., Morvan, T. and Henault, C. (2009), ‘Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model’, *Agriculture, Ecosystems & Environment* **133**(3-4), 208–222.
- Lenormand, M., Jabot, F. and Deffuant, G. (2013), ‘Adaptive approximate Bayesian computation for complex models’, *Computational Statistics* **28**(6), 2777–2796.
- Li, X., Yeluripati, J., Jones, E. O., Uchida, Y. and Hatano, R. (2015), ‘Hierarchical Bayesian calibration of nitrous oxide (N₂O) and nitrogen monoxide (NO) flux module of an agro-ecosystem model: ECOSSE’, *Ecological Modelling* **316**, 14–27.
- Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C. P. and Stumpf, M. P. H. (2014), ‘A framework for parameter estimation and modelselection from

- experimental data in systems biology using approximate Bayesian computation', *Nature Protocols* **9**(2), 439–456.
- Myrgeiotis, V., Rees, R. M., Topp, C. F. E. and Williams, M. (Submitted), 'A global parametric sensitivity analysis of the Landscape-DNDC model', *Environmental Modelling & Software* .
- Myrgeiotis, V., Williams, M., Rees, R. M., Smith, K. E., Thorman, R. E. and Topp, C. F. E. (2016), 'Model evaluation in relation to soil N₂O emissions: An algorithmic method which accounts for variability in measurements and possible time lags', *Environmental Modelling & Software* **84**(C), 251–262.
- Patil, A., Huard, D. and Fonnesbeck, C. J. (2010), 'PyMC: Bayesian Stochastic Modelling in Python.', *Journal of Statistical Software* **35**(4), 1–81.
- Rafique, R., Kumar, S., Luo, Y., Kiely, G. and Asrar, G. (2015), 'An algorithmic calibration approach to identify globally optimal parameters for constraining the DayCent model ', *Ecological Modelling* .
- Rahn, K. H., Werner, C., Kiese, R., Haas, E. and Butterbach-Bahl, K. (2012), 'Parameter-induced uncertainty quantification of soil n₂o, no and co₂ emission from hohwald spruce forest (germany) using the landscaped-ndc model', *Biogeosciences* **9**(10), 3983–3998.
- Rees, R. M., Augustin, J., Alberti, G., Ball, B. C., Boeckx, P., Cantarel, A., Castaldi, S., Chirinda, N., Chojnicki, B., Giebels, M., Gordon, H., Grosz, B., Horvath, L., Juszczak, R., Klemedtsson, Å. K., Klemedtsson,

- L., Medinets, S., Machon, A., Mapanda, F., Nyamangara, J., Olesen, J. E., Reay, D. S., Sanchez, L., Cobena, A. S., Smith, K. A., Sowerby, A., Sommer, M., Soussana, J. F., Stenberg, M., Topp, C. F. E., van Cleemput, O., Vallejo, A., Watson, C. A. and Wuta, M. (2013), ‘Nitrous oxide emissions from European agriculture – an analysis of variability and drivers of emissions from field experiments’, *Biogeosciences* **10**(4), 2671–2682.
- Saggar, S., Harvey, M., Singh, J., Giltrap, D., Pattey, E., Bromley, T., Martin, R., Dow, D., Moss, R. and McMillan, A. (2010), ‘Chambers, micrometeorological measurements, and the New Zealand Denitrification–Decomposition model for nitrous oxide emission estimates from an irrigated dairy-grazed pasture’, *Journal of Integrative Environmental Sciences* **7**(sup1), 61–70.
- Sarrazin, F., Pianosi, F. and Wagener, T. (2016), ‘Global Sensitivity Analysis of environmental models: Convergence and validation’, *Environmental Modelling & Software* **79**(C), 135–152.
- Smith, K. A., Dobbie, K. E., Thorman, R., Watson, C. J., Chadwick, D. R., Yamulki, S. and Ball, B. C. (2012), Work Package 2: The effect of fertiliser forms on nitrous oxide emissions. Component report for Defra Project NT2605 (CSA 6579), Technical report.
- Stange, C. F. and Neue, H. U. (2009), ‘Measuring and modelling seasonal variation of gross nitrification rates in response to long-term fertilisation’, *Biogeosciences* **6**(1), 2181–2192.

Sylvester-Bradley, R., Thorman, R., Kindred, D. R., Wynn, S. C., Smith, K. E., Rees, R. M., Topp, C. F. E., Pappa, V. A., Mortimer, N. D., Misselbrook, T. H., Gilhespy, S., Cardenas, L. M., Chauhan, M., Bennet, G., Malkin, S. and Munro, D. G. (2015), Minimising nitrous oxide intensities of arable crop products (MIN-NO) , Technical report.

Thorman, R. E., Kingston, E., Denton, P., Chambers, B., Dampney, P. and Williams, J. R. (2017), The Effect of N Fertiliser forms on N₂O emissions from UK arable and grassland. Experimental site in Norfolk, 2004. Version:1. [dataset] Freshwater Biological Association [publisher]. doi:10.17865/ghgno721, Technical report.

van Oijen, M., Cameron, D. R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P. E., Kiese, R., Rahn, K. H., Werner, C. and Yeluripati, J. B. (2011), ‘A Bayesian framework for model calibration, comparison and analysis: application to four models for the biogeochemistry of a Norway spruce forest’, *Agricultural and Forest Meteorology* **151**(12), 1609–1621.

van Oijen, M., Reyer, C., Bohn, F. J., Cameron, D. R., Deckmyn, G., Flechsig, M., Härkönen, S., Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkelä, A., Mette, T., Minunno, F. and Rammer, W. (2013), ‘Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe’, *Forest Ecology and Management* **289**(C), 255–268.

Van Oijen, M., Rougier, J. and Smith, R. (2005), ‘Bayesian calibration of

process-based forest models: bridging the gap between models and data.’,
Tree physiology **25**(7), 915–927.

Wainwright, J. and Mulligan, M. (2005), *Environmental Modelling: Finding Simplicity in Complexity*, Wiley.

Wang, G. and Chen, S. (2012), ‘A review on parameterization and uncertainty in modeling greenhouse gas emissions from soil’, *Geoderma* **170**(C), 206–216.